MS924D Spreadsheet Modelling & Demand Forecasting

Lab Exercises:
Holt-Winters Method for Trend and Seasonality

The exercises on this sheet allow you to extend you suite of exponential smoothing methods to include both trend and seasonal data patterns. First we shall set up a basic static model where the parameters are estimated from past data to provide projected forecasts. Second we shall show how an updating mechanism can be built in to the model so that the parameter estimates can be revised as they learn from the data.

Details about the new modelling methods are described in the relevant parts of the exercise questions below. Fuller details about each method can be found in the directed reading material. The exercises are designed to help you develop your understanding of the methods and to translate your understanding through the spreadsheet application of the methods to data. The spreadsheet is simply being used as calculator hence our focus is upon the correctness of the method implementation.

All exercises assume a time series is the data input to modelling and, in general, is denoted by $Y_1, Y_2 \ldots Y_t$ where Y_t is the observed demand at time period t. Our goal is to obtain a forecast for the next time periods, $F_{t+1}, F_{t+2}, F_{t+3}, \ldots$.

The data sets with the observed the time series are given in excel files as directed in each exercise. As well as showing your calculations in your spreadsheet, please note your interpretation of your findings as this will provide a useful record of your analysis and will help you prepare for the assignment.

Exercise 1

The managing director of an umbrella factory wants to forecast demand for his umbrellas to aid in planning production. He is aware of the seasonality of the demand for his product, and so splits the year into 5 equal periods for planning purposes (the factory is closed for 2 weeks in the summer). He has the following data on demand over the past 3 years:

Season / Year	1 Weeks 1-10 winter	2 11-20 spring	3 23-32 summer	4 33-42 autumn	5 43-52 winter
2014	129	177	34	70	130
2015	152	225	40	82	163
2016	200	252	48	104	186

(in hundreds)

(a) Plot the time series and comment on the pattern of umbrella sales.

(b) We want to set up a basic Holt-Winter's forecast model to forecast sales for 2017 and 2018. Since our data set shows a systematic increasing linear trend and also regular periodicity then we need a model that captures these patterns. Our Holt-Winters model can be written as follows

$$F_{t+m} = \left(L_t + mT_t\right) I_{t+m-l}$$

where

$F_{t+m} = m-$step ahead forecast

$L_t =$ level of underlying stationary series at time t

$T_t =$ trend at time t

$I_t =$ seasonal index at time t

$l =$ length of season in time periods

Does the expression of the model make intuitive sense? You should be able to see the linear trend model as used in Holt's method that is then multiplied by a parameter that represents the effect of season. This parameter effectively inflates or deflates the underlying trend of the data to take account of the peaks and troughs of demand associated with systematic seasonal patterns. For our data, l = 5 because there are 5 periods (or seasons) in our data. Since we want to forecast for the next two years, then m is set to 1 to 10 corresponding to the five periods in each of 2017 and 2018.

Here we set up the basic model which means that we use the observed data for 2014-2016 to estimate all parameters our model associated with the level, trend and seasonal effects. Then we use this model to project demand in future time periods.

The following steps lead you through the set-up of this model in a spreadsheet. The steps involve estimating the parameters of the modelling components in turn. Then using the model components to project the trend forward and adjust by the relevant seasonal index to obtain a forecast for a particular period.

Most calculations will involve adding columns to extend the workings across the rows corresponding to the time horizon to be modelled (e.g. 2014 period 1 – 2018 period 5). However we shall also create records of the common parameters estimated across all the parameters. So think about creating a reserved, say, at the top of your spreadsheet or to the side that is kept for the parameter estimates calculated across all the historical data (i.e. 2014-6) and will be referenced when we project forecasts for 2017-8.

i. First we need to obtain the underlying trend of the data. We start to characterise this by fitting an appropriate moving average to smooth fluctuations in the data. Since we have 5 periods in our "season" then we use an MA(5) to smooth the data. Since this moving average is used to smooth the data, we record the computed value of the moving average in the cell aligned with the centre of the observations being averaged. That is, if we average over time periods 1 to 5 then the computed value of the moving average will be recorded in a cell aligned with time period 3. On completing this step you should have a new column with the MA(5) values aligned with the mid-point of the set of values from which the average has been calculated.

NOTE that this is different from the approach we take when making a forecast with moving averages because then we usually record the forecast value in the cell corresponding to the time period to which the forecast relates. For example, if we use a moving average of time periods 1-5 to forecast the demand in time period 5 then the value is recorded in a cell aligned with time period 6. This was the approach we used earlier when we were forecasting <u>directly</u> with moving averages.

ii. Now we want to obtain the slope parameter. First create a new column and record the calculated difference between consecutive moving average values. Second compute the average of these differences and record this value in cell that is clearly labelled in your reserved space for parameters as we shall reference it in later calculations since the average difference provides an initial estimate of the slope of the linear trend. Does this operation make sense? We are effectively averaging out the local slopes between consecutive times to get an estimate of the global slope for all the past data. This global slope becomes the estimate of the trend parameter.

iii. Now we want to set up a new column that contains the values of the trend at each time period, past and future. We do this by setting the mid-point of the historical trend value to be equal to the average of the smoothed moving average values. To do this you should identify the row aligned with the middle point of the past data set (i.e. period 3 in 2015) and then record the average the smoothed MA(5) values in this cell. Then the rest of the trend values can be extrapolated by or adding the slope value (the value recorded in the cell with the trend parameter saved at the end of step ii) for time periods before or after the mid-point, respectively.

iv. Now we estimate the seasonal effects for all past observed data. In a new column calculate the seasonal value as the ratio of the appropriate observed demand (input sales data) relative to the trend values (computed in step iii) for each time period in each year (from period 1 2014 to period 5 2016). By calculating the seasonal value as a ratio we are assuming a multiplicative seasonal effect.

v. Now we obtain seasonal indices for each “season” by averaging over the seasonal values for the equivalent “season” in each year in the initialisation data set. You should obtain five seasonal indices, one for each of season 1-5. Hence for period 1, you will be averaging over the three seasonal ratios computed in step iv for 2015. 2015, 2016 to obtain one period 1 seasonal index. Record and label these five seasonal indices appropriately in your reserved space for parameter estimates.

vi. We have now obtained all the parameter estimates for the past data and are ready to start making our forecasts. We start by project the trend for each period in 2017 and 2018. That is, extend the column created in step iii to record the trend for each of the 5 periods in each of these two years. You can conduct the calculations in several ways. The easiest is to simply add on the value of the trend parameter (created in step ii) to every row for the time period 1 2017 through to period 5 in 2018 as an extension of the column of past trends created in step iii. This is a practical way of implementing the linear trend expression that is at the heart of the Holt-Winters model.

vii. Now we need to adjust the projected trend to take account of seasonal effects by multiplying by the appropriate seasonal index. That is, we record our forecasts in a new column and compute them as the product of the projected trend values from step vi times the relevant seasonal index computed in step v. For example, when forecasting for period 2 in 2017 and 2018 the appropriate trend values will be multiplied by the seasonal index computed for period 2.

viii. Plot your time series including both the observed data and the new forecasts you have just created. Does the pattern in the forecast data make sense? If you have completed your calculations correctly then the patterns for the forecast should just be an extension of the patterns observed for the relevant time periods in the past data.

ix. Now consider the situation where we have data for the actual demand for umbrellas in 2017. The data are:

Season	1	2	3	4	5
2017	209	320	60	117	219

Calculate the forecast errors for 2017 and calculate the ME, MAD, MSE, MPE and the MAPE. Keep a record of your forecast error summaries clearly labelled as we shall use them later when we compare this basic Holt-Winters forecast model with a more dynamic version.

(c) Now we shall show how to create a more dynamic version of Holt-Winters that involves updating the parameter estimates as new data become available. We shall assume that the model has been initialised using past data as in part (b) and then consider how we might use the data for 2017 to update the parameter estimates and make forecasts based on the most recent data. The data we shall use for the updating is shown below. Please imagine that this data is becoming available on a period by period basis as 2017 unfolds so that you start to appreciate the time dynamics of the data generating process.

Season	1	2	3	4	5
2017	209	320	60	117	219

We now want to dynamically update the forecasts using new data as it becomes available. That is, at the end of period 1 of 2017 a forecast is to be made to period 2, and so on.

We have an equation to update each of the level, trend and seasonal parameters in our Holt-Winters model, respectively given by

$$L_t = \alpha\left(\frac{Y_t}{I_{t-l}}\right) + (1-\alpha)(L_{t-1} + T_{t-1})$$

$$T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1}$$

$$I_t = \gamma\left(\frac{Y_t}{L_t}\right) + (1-\gamma)I_{t-l}$$

where α, β, γ are smoothing coefficients taking values between 0 and 1.

You should be able to observe that the updating formulae for each of the Holt-Winters parameters for the stationary, trend and seasonal parameters are based upon an exponential smoothing mechanism. All formula essentially take a weighted average of the last value of the parameter of interest with the most recent relevant numerical estimate of that parameter from the latest observed demand data. You should be able to see that the structure of the three formulae are the same and in the form of an "exponential smoothing", although the equations differ for each of the parameters because they relate to different components of the data patterns. The notation used is as defined in part (b).

i. You might find it useful to try writing out these equations in words (e.g. T = trend, I = seasonal index, Y = observed demand etc) if this helps you better appreciate the reasoning underpinning the updating mechanism.

ii. Think about how you might organise your spreadsheet. For example, it might make sense to create a new worksheet where include a copy of your working for steps I – v in part (b) since these correspond to the initialisation of the model parameters from the past data. You can then extend the number of rows to allow for the forecasting in 2017 and add extra columns for additional workings associated with each of the parameter updates required for this dynamic updating version.

iii. Assume the following values for your smoothing coefficients $\alpha = \beta = 0.2, \gamma = 0.6$. You might want to set up your spreadsheet so that these values can be changed so that you can try out different values to see the effect. As usual, values of the coefficients that are closer to 1 implies more responsiveness of the parameter estimates to recent data.

iv. Set up your spreadsheet to calculate updated values of the stationary, trend and seasonal indices for each period in 2017, and hence obtain forecasts for periods 1-5 in 2017.

Note 1 - for the stationary and trend parameters you will need to update based on the value from the previous time period (e.g. period 1 in 2017 will be an update of period 5 in 2016). However for the seasonal indices then the updates need to correspond to the appropriate season. (e.g. period 3 in 2017 will be an update of period 3 in 2016).

Note 2 – in our basic Holt-Winters method set up in part (b) we did not need to explicitly make calculations using the stationary value (L). This was because our operations in step vi in part (b) projected the trend for 2017 and 2018 from the level in period 5 in 2016 quite naturally by adding on the trend parameter estimate to the last value. For our dynamic version of Holt-Winters with updating we need to be more explicit in referencing the cell that contains the last value of the level which we shall update. Please use the value of the cell corresponding to period 5 of 2016 calculated in step iii of part (b) as your value for L at t-1, assuming t corresponds to period 1 in 2017.

v. Plot your forecasts and the past data. Do you forecasts make sense in relation to the patterns of past demand?

vi. Since you have the data for 2017, compute the error between the forecasts generated under the dynamic updating and the observed demand. Also, obtain the usual set of forecast error summaries (e.g. MAPE etc).

(d) Compare your forecasts for 2017 with obtained in parts (b) and (c) based on your time plots and your set of forecast error calculations. Which do you think is most useful forecasting method in an operational context and why?

(e) If you have set up your spreadsheet in a way that allows you to change the specified values of the smoothing coefficients, then try out different combinations of coefficient values. Think through what effect you might expect this to have on your model and then examine the effects on the forecasts generated and the errors statistics. What set of smoothing values would you recommend and why?

Exercise 2

The table below, also given in the Excel file "Air Passengers", shows the number of inward passengers to the UK travelling by air on scheduled services, during the period from the first quarter of 1995 to the third quarter of 2004.

We shall use this data set to show how we can generate de-seasonalised values under an additive model so that we understand the underlying data patterns adjusted appropriately for time of year effects. We will then extend the analysis to obtain forecasts for 2014.

Less direction is given in each step since we are creating a variant of the process followed in Exercise 1. The variants relate to, for example, the length of the season, the form of the seasonal effect, and the interim purpose of modelling which is to de-seasonalise as well as forecast.

Year	*Quarter*	*Passengers*	*Year*	*Quarter*	*Passengers*
1995	I	9337	2000	I	13665
1995	II	13335	2000	II	18636
1995	III	16545	2000	III	22743
1995	IV	11287	2000	IV	16034
1996	I	10262	2001	I	14302
1996	II	13572	2001	II	17782
1996	III	16583	2001	III	22987
1996	IV	12075	2001	IV	14684
1997	I	11003	2002	I	14269
1997	II	14926	2002	II	18858
1997	III	18147	2002	III	23347
1997	IV	13066	2002	IV	16702
1998	I	11929	2003	I	15150
1998	II	16323	2003	II	19430
1998	III	19949	2003	III	24521
1998	IV	14251	2003	IV	17818
1999	I	13083	2004	I	16447
1999	II	17249	2004	II	21602
1999	III	21137	2004	III	26213
1999	IV	15110			

Source: Monthly Digest of Statistics.

(a) Plot the time series and comment upon the patterns displayed in the number of air passengers.

(b) Model the trend in the data using an appropriate moving average. We now have an even number of seasons per year and so a form of centred moving averages are appropriate. Centred moving averages means calculating the moving averages twice so that we centre the averaged value in the middle of the data being averaged.

(c) Plot the smoothed data set created in part (b) on the original observations and comment on how well the smoothed data tracks the trend.

(c) Compute the seasonal values for each quarter by computing the difference (or deviation) between the observed and the trend values, and averaging for each season to obtain the four seasonal indices. Differencing corresponds to an additive time series model (unlike a multiplicative time series model where the seasonal components are calculated as the ratio of observed to trend as in Exercise 1). In general, an additive model is more appropriate when the amplitude of seasonal cycles are the same, while a multiplicative model is more appropriate the amplitude of seasonal cycles may vary.

(d) Compute the seasonally adjusted time series values by subtracting the appropriate seasonal indices from the actual observations for each time point. We shall call these new values the de-seasonalised series.

(e) Plot the de-seasonalised series and note the patterns you observe? Does it match what you expect after you have removed the seasonal effects?

(e) Using a basic additive Holt-Winters make forecasts for the last quarter of 2004 by using an appropriate selection of the past data. Justify your choice of data to initialise your forecasting model. You will need to compute the trend in a similar manner to Exercise 1 and then project this trend and adjust by adding the value of the appropriate seasonal index.